Tema 6: Introducción al muestreo. Estimadores

VARIABLE Cierta variable aleatoria X se distribuye según la función de densidad:

$$f(x,\theta)=\frac{\theta^2}{2x};\quad 0\le x\le \frac{2}{\theta},$$

siendo E(X)=θ y Var(X)=1-θ².

Para estimar el parámetro θ, se ha tomado una muestra aleatoria simple de tamaño n proponiéndose el estimador: $T_1=\overline{X}+3$.

a) Obtener e interpretar el sesgo del estimador
b) ¿Cuál es la varianza del estimador?
c) Además se propone otro estimador T_2, insesgado y con varianza $\frac{1-\theta^2}{n}$. ¿Cuál de los dos estimadores es más eficiente para estimar θ?
d) ¿Cuál es la estimación obtenida si en una muestra aleatoria se ha observado una media de 3,5?

Resultados

a) $B_T(\theta)=3$

b) Var(T)= $\frac{1-\theta^2}{n}$

c) Es más eficiente el estimador T_2

d) $\hat{\theta}=6{,}5$

ESTIMADORES DE LA MEDIA Para estimar el parámetro μ de una población se toma una muestra aleatoria simple de tamaño 4, proponiéndose los dos estimadores siguientes:

$$T_1=\overline{X}_4 \qquad\qquad T_2=\frac{X_1+2X_3+X_4}{4}$$

¿Cuál de los dos estimadores es más eficiente?

Resultados

$$ECM_{T_1}(\mu)=\frac{\sigma^2}{4}$$

$$ECM_{T_2}(\mu) = \frac{3\sigma^2}{8}$$

El estimador T_1 es más eficiente para estimar μ que el estimador T_2.

UNIVERSITARIOS Se desea conocer la proporción de universitarios entre la población menor de 25 años de determinado país (p). Algunos estudios señalan que dicha proporción es del 25% mientras otros aseguran que es del 30%.

Seleccionados aleatoriamente 20 jóvenes del colectivo estudiado, 4 manifestaron ser universitarios. Con esta información, justificar cuál es la estimación más verosímil del parámetro p.

Resultados

p=0,25 es la estimación más verosímil.

SALTOS El número de saltos nulos realizados por un atleta de alta competición en un mes es aleatorio, y su función de probabilidad viene dada por la siguiente expresión:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, ...$$

a) Obtener la función de verosimilitud para una muestra aleatoria simple de tamaño n.
b) Deducir el estimador máximo verosímil de λ. Si en una muestra concreta se ha observado una media de 1,75 saltos nulos por mes, ¿cuál es la estimación máximo verosímil de λ?
c) Obtener un estimador de λ por el método de los momentos.

Resultados

a)

$$L(x_1,...,x_n,\lambda) = \prod_{i=1}^{n} P(x_i,\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

b)

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{\lambda} = \overline{x} = 1,75$$

c)
$\hat{\lambda} = \overline{x}$

AYUNTAMIENTO **Un ayuntamiento está estudiando la duración de las bombillas instaladas en las farolas de su alumbrado público, que es una variable aleatoria X con:**

$$f(x) = \frac{1}{\theta} e^{-\frac{(x-10)}{\theta}} \qquad x > 10; \theta > 0$$

E(X)= θ +10 ; Var(X)=θ^2
Para estimar el parámetro θ se decide observar la duración de una muestra aleatoria de 10 bombillas.
a) Deducir la expresión del estimador de máxima verosimilitud de θ.
b) Analizar si el estimador obtenido es insesgado y calcular su error cuadrático medio.

Resultados

a) $\hat{\theta} = \overline{X} - 10$
b) El estimador es insesgado.

$$\mathrm{ECM}_{\hat{\theta}}(\theta) = \frac{\theta^2}{n}$$

ESTIMADORES

Con el objetivo de aproximar el parámetro θ, $(\theta > 0)$, se ha seleccionado una m.a.s. de tamaño 4 y se consideran dos estimadores sobre los que se dispone de la siguiente información:

Estimador	Esperanza	Varianza
T_A	$E[T_A] = \theta$	$\mathrm{Var}[T_A] = \frac{\theta}{4}$
T_B	$E[T_B] = \frac{5\theta}{4}$	$\mathrm{Var}[T_B] = \frac{15\theta}{16}$

a) Obtener e interpretar el sesgo de los dos estimadores propuestos.
b) ¿Cuál de las dos expresiones es más eficiente para estimar θ?

Resultados

a) $B_{T_A}(\theta) = 0$

$$B_{T_B}(\theta) = \frac{\theta}{4}$$

b) $\mathrm{ECM}_{T_A}(\theta) = \frac{\theta}{4}$

$$ECM_{T_B}(\theta) = \frac{\theta^2}{16} + \frac{15}{16}\theta = \frac{\theta}{4}\left(\frac{\theta + 15}{4}\right)$$

El estimador T_A es más eficiente para estimar θ.

BECAS Un organismo que ha convocado becas juveniles se plantea estudiar el gasto mensual de sus potenciales beneficiarios. Gracias a una encuesta se ha accedido a información relativa a los gastos de 4 jóvenes representativos del colectivo estudiado.

Justificar cuál de las siguientes expresiones es más eficiente para estimar el gasto mensual esperado:

$$T_a = \sum_{i=1}^{4} \frac{X_i}{4}$$

$$T_b = \frac{X_1 + 0{,}5X_2 - 0{,}5X_3 + 3X_4}{3}$$

Resultados

$$ECM_{T_a}(\mu) = \frac{\sigma^2}{4}$$

$$ECM_{T_b}(\mu) = \left(\frac{\mu}{3}\right)^2 + \frac{10{,}5\sigma^2}{9}$$

El estimador T_a es más eficiente para estimar el gasto mensual esperado.

RESTAURANTELos tiempos de espera (en minutos) de los clientes de un restaurante hasta ocupar una mesa se distribuyen uniformemente en el intervalo [0,θ].
a) Estudiar el sesgo de la media muestral como estimador del parámetro θ.
b) Obtener un estimador de θ por el método de los momentos.
c) Comparar la eficiencia de los dos estimadores anteriores de θ.

[Extraído de *Análisis de Datos Económicos II. Métodos Inferenciales*, problema 5.5, pág. 299]

Resultados

a) $B_{\overline{X}}(\theta) = -\frac{\theta}{2}$

b) $\hat{\theta} = 2\overline{X}$

c)

$$ECM_{\overline{X}}(\theta) = \frac{\theta^2(3n+1)}{12n}$$

$$ECM_{2\overline{X}}(\theta) = \frac{4\theta^2}{12n}$$

El estimador $2\overline{X}$ es más eficiente que $\overline{X}$ para estimar θ.

SUPERMERCADO **Un supermercado va a realizar un estudio sobre la proporción de clientes interesados en la ampliación de su horario comercial, para lo cual se investigará una m.a.s. de clientes.**

a) Justificar cómo se definirán las variables muestrales, estudiando su distribución de probabilidad.
b) Deducir la expresión de la función de verosimilitud muestral. ¿Qué significado tiene?
c) Supongamos que la muestra es de tamaño 5 y consideremos el estadístico T definido como:

$$T = \frac{2X_1 + X_2 - X_3 + X_4 + 2X_5 + 10}{5}$$

Comprobar si T es un estimador insesgado de la proporción de clientes interesados en la ampliación de horario y calcular su error cuadrático medio.

Resultados

a) $X_i \approx B(p)$

b) $L(x_1, x_2, \ldots x_n, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$

c) $E(T) \neq p$

$$ECM_T(p) = 4 + \frac{11pq}{25}$$

LLAMADAS **La duración en minutos de las llamadas telefónicas realizadas en una empresa es una variable aleatoria X con:**

$$f(x) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \quad \text{para} \quad x > 0; \quad \theta > 0 \quad \text{siendo} \quad E(X) = \theta$$

a) Obtener el estimador máximo verosímil del parámetro θ.
b) Analizar el sesgo del estimador obtenido.
c) Obtener un estimador de θ por el método de los momentos.

Resultados

a) $\hat{\theta} = \overline{X}$

b) $B_{\hat{\theta}}(\theta) = 0$

c) $\hat{\theta} = \overline{X}$